

# Loughborough University Institutional Repository

---

## *Using social media to analyse the Scottish Referendum journey*

This item was submitted to Loughborough University's Institutional Repository by the/an author.

**Citation:** JACKSON, T.W. and SYKORA, M.D., 2015. Using social media to analyse the Scottish Referendum journey. Software Quality Management, Loughborough University, 30th March 2015

**Additional Information:**

- This is a conference paper.

**Metadata Record:** <https://dspace.lboro.ac.uk/2134/17968>

**Version:** Accepted for publication

**Publisher:** Loughborough University

**Rights:** This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the published version.

# Using Social Media to Analyse the Scottish Independence Referendum Journey

T. W. Jackson, M. D. Sykora, and S. Elayan  
Centre for Information Management, Loughborough University,  
Loughborough, Leicestershire LE11 3TU, UK

[T.W.Jackson@lboro.ac.uk](mailto:T.W.Jackson@lboro.ac.uk); [M.D.Sykora@lboro.ac.uk](mailto:M.D.Sykora@lboro.ac.uk); [S.Elayan2@lboro.ac.uk](mailto:S.Elayan2@lboro.ac.uk)

## Abstract

The effective management and efficient use of data is a major challenge for government, academia and industry. The scale, diversity and distributed nature of current and emerging data assets are increasing. For example, every sixty seconds there are 433,000 tweets, 2.66 million Google searches, 4.7 million Tumblr posts, and 139 million emails sent. As data becomes ever more ubiquitous and critical to decision making it is vital that it is turned into meaningful information. The aim of the research discussed in this paper was to determine the usefulness of the EMOTIVE system, developed by the authors, to track the Scottish Independence Referendum and if it could be used to track the general public emotions towards the UK 2015 General Election by creating a new mobile phone application called VOTEBEE (Voter Opinion Tracked by EMOTIVE Engine). In this paper we employ an ontology engineering approach to the problem of fine-grained emotion detection in sparse messages. Messages are also processed using a custom NLP pipeline, which is appropriate for the sparse and informal nature of text encountered on micro-blogs or tweets. Our approach detects a range of eight high-level emotions; anger, confusion, disgust, fear, happiness, sadness, shame and surprise. The paper analyses the Scottish Independence Referendum data using EMOTIVE to determine what can we learn from monitoring social media networks and if there is a requirement to develop systems like VOTEBEE to help determine voter interest and increase debates in the lead up to the 2015 UK General Election.

## **1.0 Introduction**

With the uptake of social media, such as Facebook and Twitter, there is now a vast amount of new user generated content on a daily basis, for example, every sixty seconds there are 433,000 tweets [1]. Much of content is in the form of short, informal free-form text. Businesses, institutions, governments and law enforcement organisations are now actively seeking ways to monitor and more generally analyse public response to various events, products and services.

Due to the real-time nature and the value of social media content for monitoring entities and events of significance, automated sentiment analysis and semantic enrichment techniques for social media streams have received considerable attention in the literature. These techniques are central to monitoring social-media content, which is now becoming a significant business with commercial, institutional, governmental and law enforcement interest into its applications. Prior work in sentiment analysis especially has focused mostly on negative-positive sentiment classification tasks. Although numerous approaches employ highly elaborate and effective techniques with some success, the sentiment or emotion granularity is generally limiting and arguably not always most appropriate for real-world problems. In this paper a newly developed ontology based system is employed, to semantically enrich Tweets with fine-grained emotional states, in order to analyse the subjective public reactions to a wide selection of recent events. The approach detects a range of eight high-level emotions and their perceived strength (also known as activation level), specifically; anger, confusion, disgust, fear, happiness, sadness, shame and surprise. A set of emotional profiles for different events is obtained and an in-depth analysis of the emotional responses is presented. The feasibility and potential benefits of automated fine-grained emotional event response analysis from social-media is illustrated and how it will be used for the General Election 2015 will be discussed.

The remainder of the paper is organised as follows. Section 2 introduces some background and prior work in the sentiment analysis field, and gives brief method details. The research design is presented in section 3. Section 4 analyses and discusses the data collected. There are potentially many ethical dilemmas around

using open data sources and the impact this could have on society is covered in section 6. The paper is concluded in section 7.

## **2.0 Background**

Recently there has been a lot of attention given to automated sentiment analysis and semantic enrichment (e.g. geo-location inference, named entity recognition, topic classification, etc.) of social media text streams, such as Tweets and Facebook status updates. Largely motivated by the insights and value that such datasets were shown to provide [2,3,4,5,6], university researchers and organisations have started to invest large sums of money into this area. It has also been evidenced that during times of natural crises and terrorist incidents Twitter is often the first medium through which the news breaks and through which individuals express their initial impressions and emotions relating to the events [7 to 11]. Social-media streams, in general, allow for observing large numbers of spontaneous, real-time interactions and varied expression of opinion, which are often fleeting and private [12]. Miller [12] further points out that some social scientists now see an unprecedented opportunity to study human communication, which has been an obstacle up until recently. O'Connor et al. [13] demonstrated how large-scale trends can be captured from Twitter messages, based on simple sentiment word frequency measures. The researchers evaluated and correlated their Twitter samples against several consumer confidence and political opinion surveys in order to validate the approach, and have pointed out the potential of social-media as a rudimentary yet powerful polling and survey methodology. Motivated by such work, this paper will specifically focus on automated fine-grained emotion analysis (also known as advanced sentiment analysis) for a similar event to the UK General Election, the Scottish Independence Referendum in 2014. The study will look at the evoked emotions around the Scottish Independence Referendum and how the cut down version of EMOTIVE, VOTEBEE, could be used in the UK General Election.

## **3.0 Research Design**

An in-depth overview of prior academic work in the sentiment analysis field is provided in Thelwall et al. [14]. The approach used in this paper [15] broadly falls under the lexicon / linguistic analysis approach, from the three approaches presented in Thelwall et al. [14] – except that we draw on emotion terms from within an ontology with a richer semantic representation than commonly used emotion term-lexicons. Although numerous approaches employ highly elaborate and effective techniques with some success, the sentiment or emotion granularity is generally limiting. Specifically, there are three main problems with existing approaches. 1-Notions of affect and sentiment have been rather simplified in current state-of-the-art, often confined to their assumed overall polarity (i.e. positive / negative), Thelwall et al. [14]. 2-Another problem with polarity-centric sentiment classifiers is that they generally encompass a vague notion of polarity that bundles together emotion, states and opinion [16]. 3-There is no common agreement about which features are the most relevant in the definition of an emotion and which are the relevant emotions and their names [17]. In the emotion analysis employed in this paper, sentiment is fine-grained, based on the widely accepted Ekman’s emotions [18] from social psychology, while other work on emotions was also considered [18 – 20] and is further discussed in Sykora et al. [15]. Only explicit expressions of emotions are extracted, and ambiguous emotional expressions, such as certain moods and states that are not expressing emotions are ignored on purpose, as opposed to Bollen et al., [16], Lansdall-Welfare et al. [5], and Choudhury and Counts [22]. The EMOTIVE ontology employed in this paper was designed to detect a wider range of well recognised human emotions, such as ‘surprise’, ‘disgust’, or ‘confusion’, but at the same time differentiate emotions by strength (e.g. ‘uneasy’, ‘fearful’, ‘petrified’). In addition to the basic emotions, the ontology also covers and handles negations, intensifiers, conjunctions, interjections, and contains information on the perceived strength (also known as activation level) of individual emotions, whether individual terms and phrases are slang or used in standard English and their associated POS (Parts-of-Speech) tags, where this aids to resolve ambiguity. In Sykora et al. [15] our technique was evaluated and compared to Choudhury and Counts [22] and Thelwall et al. [14] – SentiStrength 2 – in terms of emotion detection and emotion strength scoring, respectively. Good results, comparable with state-of-the-art were

achieved and a high f-measure for emotion extraction on an initial test dataset was reported.

The datasets analysed within this paper were continuously retrieved from Twitter, using the standard REST Twitter Search API. The retrieval occurred during the related time-period of an event and a search-term or hashtag, known to be extensively used by the Twitter community for that event was chosen by a microblogging expert. For most events of interest data collection would occur during the days / time-period of the event, or the days immediately following the event in order to collect the related reactions, chatter and emotions. Often the selected term or hashtag used for the data collection would also be trending, i.e. according to Twitter trends. The maximum possible number of tweets, given the API limitations and compatible with Twitter's terms of service, was automatically retrieved using custom developed scripts. In total over 45,000 tweets were collected and analysed for the Scottish Independence Referendum results provided in section 4.

#### **4.0 Results – Scottish Independence Referendum 2014**

The Scottish Independence Referendum was a referendum on Scottish independence that took place in Scotland on 18 September 2014. The independence referendum question, which voters answered with "Yes" or "No", was "Should Scotland be an independent country?". The "No" side won, with 2,001,926 (55.3%) voting against independence and 1,617,989 (44.7%) voting in favour. The turnout of 84.6% was the highest recorded for an election or referendum in the United Kingdom since the introduction of universal suffrage.

During the run up to the vote for independence the EMOTIVE system tracked tweets to do with the Scottish Independence Referendum. Table 1 shows the size of the emotions for the 10, 11, 12, and 14<sup>th</sup> of September. During this time there was 24 hour rolling coverage of the Scottish Independence Referendum, with reaction and analysis covered across all media. All of these emotions are linked to events that occurred during the campaign. For example, when the UK Prime Minister, David Cameron, gave a speech on the 10<sup>th</sup> of September about the referendum and

said “Please don’t go” many thought this was a sign of desperation and a step to far to keep the Scottish part of the Kingdom. This lead to an overall rise in the feeling of sadness (as Shown by Figure 1) and a realisation that the United Kingdom might cease to exist in its current form.

<b>Data collected on 10/09/2014</b>		<b>Data from 11/09/2014</b>	
Anger	2%	Anger	9%
Confusion	3%	Confusion	3%
Disgust	12%	Disgust	18%
Fear	18%	Fear	28%
Happiness	22%	Happiness	12%
Sadness	23%	Sadness	8%
Shame	5%	Shame	6%
Surprise	15%	Surprise	15%
<b>Data collected on 12/09/2014</b>		<b>Data from 14/09/2014</b>	
Anger	4%	Anger	2%
Confusion	4%	Confusion	4%
Disgust	5%	Disgust	15%
Fear	18%	Fear	22%
Happiness	25%	Happiness	23%
Sadness	13%	Sadness	10%
Shame	9%	Shame	5%
Surprise	21%	Surprise	19%

**Table 1 – Data Collected from EMOTIVE showing emotions**

As shown by Table 1 there has always been a strong feeling of ‘Fear’. When analysing the EMOTIVE picture gallery as shown by Figure 1 it highlighted some of the issues that lead to the general public fears. The EMOTIVE gallery (shown by Figure 1) shows the images being incorporated into tweets that contain the words, hashtags or @’names’ being monitored. The uncertainty (or fear) of a ‘Yes’ or ‘No’ featured heavily in the analysis and a good example of this was surrounding the issue of pensions. If Scotland became independent would the Scots still be entitled to their pension. In Figure 1 on the left hand side, there is a red circle and a red cross which denotes a letter sent from the Pension Service, which is shown by Figure 2.

The response from the Pension Service was rather ambiguous and discussed issues like ‘being abroad’, not being part of Britain, and no increase in the payable

pension whilst living outside of Britain. This was indicative of the issues that were raising ‘fear’ amongst the general public.

What was interesting to note is during the campaign the large sums of money that must have been spent to produce images to help both the ‘Yes’ and ‘No’ campaign. New images were being posted every couple of minutes to try and win the vote. This is a good indication that both parties had now started to spend some of their marketing budget on social media to win over the voters.



Figure 1 – Sample of images posted on twitter during the campaign.

Using the EMOTIVE system during the Scottish vote could have provided the two campaigns with a great insight of what the voters were concerned about it terms of the emotions being evoked. As mentioned, Fear was a major factor in the decision on whether to vote ‘Yes’ or ‘No’, but if both campaigns were provided with the analysis provided by a tool like EMOTIVE they could have used that emotion to help their campaign. For example, the ‘No’ campaign might have wanted to play on this emotion to add uncertainty and ‘best stick with what you know’ approach. The ‘Yes’ campaign could have allayed these fears by providing evidence about the truth.



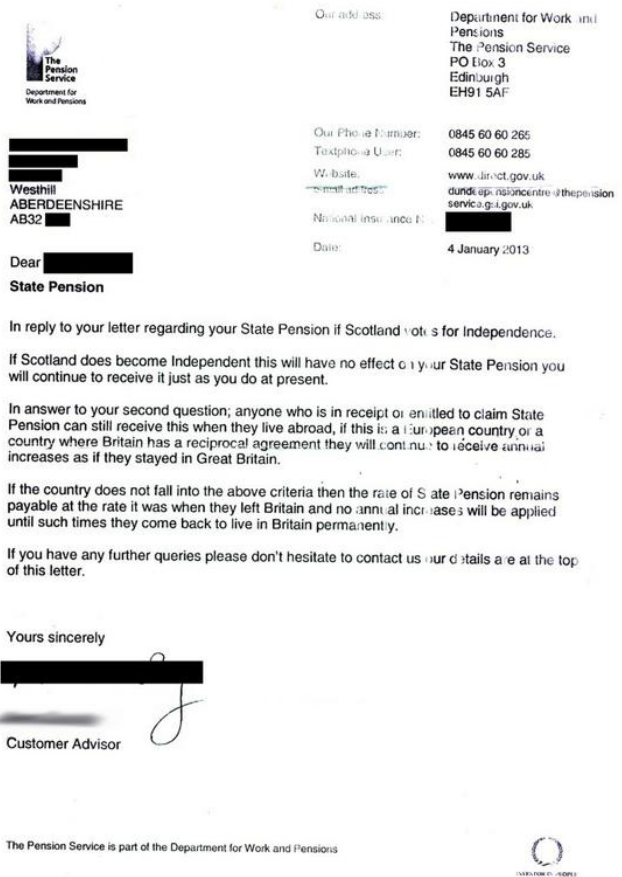


Figure 2 – Ambiguous letter from the pensions service

Throughout the campaign, the authors were asked numerous times by the media if EMOTIVE could predict the outcome of the referendum. The authors explained that at this stage the system could only highlight issues raised by potential voters which the parties could then try and address to win the vote. However, once the system has monitored a number of elections it could potentially use Case Base Reasoning (CBR). Using CBR it could determine a likely winner by analysing the flows of emotions during the campaign and look for trends that would indicate the potential winner. The EMOTIVE system is going to be used for the 2015 UK General Election and will be used to inform news outlets like 'The Conversation'. A smaller version of EMOTIVE will be created so the masses can see the ebb and

flow of general election emotions via a mobile phone application called VOTEBEE.

#### **4.1 VOTEBEE – 2015 UK General Election**

VOTEBEE is a mobile phone application that uses the EMOTIVE engine to track the emotions towards political leaders, parties, partners of the leaders and policies. It will be available for the 2015 UK General Election and it will provide an alternative view to how the election is unfolding for the various parties. As shown by the Scottish Independence Referendum analysis there is some potential in monitoring social media networks to determine the issues in real-time. The VOTEBEE application will show the strength of emotions towards each of the seven parties, party leaders, partners and policies. The user will be able to then click on one of the parties, partners, leaders or policies to receive a breakdown of the emotions towards them. The VOTEBEE system will then show how the emotions have developed over time, the top tweets and the top ten pictures. From a party viewpoint they will be able to identify what the electoral are feeling towards their campaign. Like the Scottish Independence Referendum the authors expect fear to play a strong role in the party campaigns.

With the ever increasing use of social media, with 433k tweets every 60 seconds around the globe, the live tweets can provide the ears and eyes of people on the ground. Any individual using social media is effectively playing the part of a sensor that can be used to determine the social barometer towards events, products or services. One moral dilemma is how far should scientists go when developing new systems that use multiple data sets to provide new interpretations and insights.

#### **5.0 Big Data and the Moral Dilemma?**

As data sources become readily available or at worst can be purchased at minimal cost, how far should we go in joining up these data sets to provide a bigger more detailed picture. In terms of social media, for example, all FaceBook posts, Tweets, and blogs can be purchased and analysed by anyone willing to pay a small price to obtain the data. As has been outlined in this paper and previous research (Sykora et al., 2013) the use of tweets can be used to predict riots and potentially indicate

required interventions, through to, gaining feedback from products or services that help organisations tailor their next offering. This might not seem to be a bad thing, but looking ahead in this journey might not be so appetising. It is clear to see that if Tesco Clubcard worked with Cancer Research it might be possible to determine who might get cancer by looking at their diet or the chemicals they use for cleaning. At first glance this can only be a good thing. However, if it was discovered that cheap to buy food had the ingredients that could lead to Cancer would suppliers be forced to change the ingredients which then might put up the cost of the food and make the poverty divide even greater. We have already seen IBM's Watson work with oncologists to take all the available data for the last ten years to help cancer treatment [12]. The Watson system can now provide the best course of treatment given the patients details which if scaled up to other areas of medicine could see a decline in the need for Doctors. Potentially systems could start to replace professions, which then leaves the question of what jobs will we do in the future. Although this is many years off if unchecked, this type of work outlined in the paper is the start of a system run society and it is the moral dilemma that researchers have to face when developing new techniques to help push the research boundaries forward to fully utilise big data.

## **6.0 Conclusion**

In summary the effective management and efficient use of data is a major challenge for government, academia and industry. However, it is no longer the manipulation of big data, but we have moved into the era of trying to make sense of combining existing information sets to provide a richer more detailed picture of an event, service or product. The scale, diversity and distributed nature of current and emerging data and information assets are increasing. The aim of the research was to determine the usefulness of the EMOTIVE system to track the Scottish Independence Referendum and if it could be used to track the general public emotions towards the UK 2015 General Election. The results showed that if parties monitored the social media feeds it could provide them with an insight into the issues the voters were concerned with, and issues that they would need to address. However, EMOTIVE could not determine the winner of election until it had built up a number of cases so a technique called Case Base Reasoning could be used to

potentially identify the winner of an election given the emotions evoked from the general public.

The paper concluded by highlighting the moral dilemmas of building such systems and the need for caution. There is a need for Governments and Societies, like the British Computer Society [23], to take a bigger role in shaping the emerging discipline of data analytics and to provide ethical guidelines given the fast moving nature of system development in this area.

## 7.0 References

1. Go-Globe.com, 2014, In sixty seconds, [last accessed 14/11/2014]
2. Chew, C. and Eysenbach, G. (2010) "Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak", PLOS One, Vol 5, No. 11.
3. O'Connor, B., Balasubramanyan R., Routledge B. and Smith N. (2010) "From Tweets to Polls: LinkingText Sentiment to Public Opinion Time Series", paper presented at the fourth International AAAI Conference on Weblogs and Social Media, Washington D.C., USA, May.
4. Tumasjan, A., Sprenger, T. O. and Welpe, I. M. (2010) "Predicting elections with twitter: What 140 characters reveal about political sentiment", paper presented at the fourth International AAAI Conference on Weblogs and Social Media, Washington D.C., USA, May.
5. Lansdall-Welfare, T., Lampos, V. and Cristianini, N. (2012) "Effects of the Recession on Public Mood in the UK", paper presented at the 21st International Conference companion on World Wide Web, Lyon, France, April.
6. Abel, F., Hauff, C., Houben, G., Stronkman R. and Tao, K. (2012) "Semantics + Filtering + Search = Twitcident Exploring Information in Social Web Streams", paper presented at the 23rd ACM International Conference on Hypertext and Social Media, Milwaukee, USA, June.
7. Beaumont C. (2008) "Mumbai attacks: Twitter and Flickr used to break news", The Telegraph, <http://www.telegraph.co.uk/news/worldnews/asia/india/3530640/Mumbai-attacks-Twitter-and-Flickr-used-to-break-news-Bombay-India.html> [27th November 2008].
8. Cashmore P. (2009) "Mashable: Jakarta bombings - Twitter user first on the scene", [online], Mashable, <http://mashable.com/2009/07/16/jakarta-bombings-twitter/> [16th July 2009].
9. Sakaki, T., Okazaki, M. and Matsuo, Y. (2010) "Earthquake shakes Twitter users: real-time event detection by social sensors", paper presented at the 19th International Conference on World Wide Web, Raleigh, USA, April.
10. Cheong, M. and Lee, V. C. S. (2011) "A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter", Journal of Information Systems Frontiers – Springer, Vol 13, No. 1, pp 45-59.

11. Colbaugh, R. and K. Glass, "Detecting emerging topics and trends via predictive analysis of 'meme' dynamics", Proc. 2011 AAAI Spring Symposium Series, Palo Alto, CA, March 2011. Miller, G. (2011) "Social scientists wade into the tweet stream", Science, Vol 333, No. 6051, pp 1814-1815.
12. IBM WATSON, <http://www.mskcc.org/cancer-care/watson-oncology> (last accessed 18/03/2015).
13. O'Connor, B., Balasubramanyan R., Routledge B. and Smith N. (2010) "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series", paper presented at the fourth International AAAI Conference on Weblogs and Social Media, Washington D.C., USA, May.
14. Thelwall, M., Buckley, K. and Paltoglou, G. (2012) "Sentiment Strength Detection for the Social Web", Journal of the American Society for Information Science and Technology, Vol 63, No. 1, pp 163-173.
15. Sykora, M., Jackson, T. W., O'Brien A. and Elayan, S. (2013) "EMOTIVE Ontology: Extracting fine-grained emotions from terse, informal messages", IADIS International Journal on Computer Science and Information Systems, Vol 8, No. 2, pp. 106-118.
16. Bollen, J., Mao, H. and Pepe, A. (2011) "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena", paper presented at the fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, July.
17. Grassi, M. (2009) "Developing HEO human emotions ontology", Biometric ID Management and Multimodal Communication, Springer Berlin Heidelberg, pp 244-251.
18. Ekman, P. and Richard, J. (1994) "All emotions are basic, The nature of emotion: Fundamental questions", Oxford University Press, New York, pp 15-19.
19. Plutchik, R. (1980) "Emotion: A Psychoevolutionary Synthesis", Longman Higher Education, 1980.
20. Drummond, T. (2004) "Vocabulary of Emotions", [online], North Seattle Community College, Available from <http://www.sba.pdx.edu/faculty/mblake/448/FeelingsList.pdf> [last viewed 9th January 2012].
21. Izard, C. E. (2009) "Emotion theory and research: Highlights, unanswered questions, and emerging issues", Annual Review of Psychology, Vol 60, No. 1, pp 1-25.
22. Choudhury, M. and Counts, S. (2012) "The Nature of Emotional Expression in Social Media: Measurement, Inference and Utility", Technical Report: Microsoft.
23. The British Computer Society, <http://www.bcs.org/>, last accessed 18/03/2015.